

# Evaluation Approaches

## Contents

### Module\_8 – Evaluation Approaches

8.1 - Introduction and Background . . . . .	2
8.2 - Guidelines and Standards of Evaluation . . . . .	2
8.3 - Cognitive Walkthroughs . . . . .	3
8.4 - Heuristic Evaluations . . . . .	4
8.5 - Think-Aloud Evaluations . . . . .	5
8.6 – Classic Experiments . . . . .	6
8.7 - Focus Groups . . . . .	7
8.8 - Concise Summary . . . . .	10
8.9 - Extended Resources . . . . .	11
8.10 - References . . . . .	12
8.11 - PowerPoint Presentation–Refer to D2L	

**8.1 - Introduction and Background**

Interactive systems are present in all domains and sectors. They dominate much of how we interact with data, objects, and even other people in our world today. With the dominance of interactive systems there comes a need to evaluate such systems.

Generally, there are three categories of evaluating user interfaces: controlled settings with users, natural settings with users, and settings not involving users. Typically these three categories of evaluation are combined to create a stronger evaluation of a system (Preece et. al. 2015). Testing that does not involve users can be cheaper, but including users generates a richer data set, because evaluators will be able to view people that are unfamiliar with the system and flow of events.

Within the three general categories, there are different methods for conducting an evaluation. The methods that will be discussed in detail are cognitive walkthroughs, heuristics evaluations, think-aloud evaluations, classic evaluations, and focus groups. These methods of evaluation are usually employed by a group of 3 to 5 usability experts. The usability experts find any usability concerns that exist within the iteration that they are evaluating; they then assign a number to each concern that rates the severity (Gonzalez-Holland, Whitmer, Moralez, and Mouloua, 2017).

Now we have an idea of what tools are used in the evaluation of interactive systems, but before a system can be evaluated, the following need to be established: why the system is being evaluated, what is going to be evaluated, the location in which will it be evaluated, and the time during which it will be evaluated (Preece et. al. 2015). These are all very important questions to ask before the process of evaluation can occur, as they will help determine what the evaluation of the system will look like. For example, a small app will most likely not be evaluated using the classical evaluation and tens of thousands of participants; on the other hand, a massive medical system will need more than ten participants in a focus group.

**8.2 – Guidelines and Standards of Evaluation**

The guidelines for user interfaces focus heavily on user experience. Two main design guidelines that are widely accepted are Nielsen’s ten “heuristics,” and Shneiderman and Plaisant’s eight golden rules.

<b>Jakob Nielsen’s 10 heuristics</b>	<b>Shneiderman &amp; Plaisant’s 8 Golden Rules</b>
1. Flexibility and Efficiency of Use	1. Strive for Consistency
2. Standards and Consistency	2. Offer informative feedback
3. System Status Visibility	3. Allow users to save and use shortcuts
4. Error Prevention for Users	4. Design dialogues to yield closure

5. Reduce User's Memory Load	5. Offer simple error handling
6. Help and Guidance	6. Permit simple reversal of actions
7. Minimalist design and emphasis on Aesthetic	7. Support internal locus of control
8. User Control and Freedom	8. Reduce short-term memory load
9. Help User recognize, diagnose, and recover from errors	
10. A match between the real world and the system	

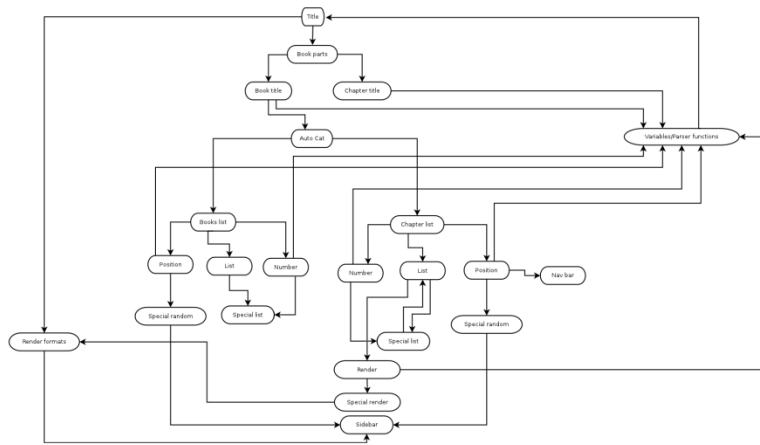
A heuristic is, in simple terms, an approach to solving a problem. It is a trial-and-error method where investigation and problem-solving is encouraged. The heuristics above do not serve to tell evaluators precisely how to achieve their goals, but rather the guidelines help to produce the greatest outcome when followed. These different guidelines can be combined as well (see “Examination of the use of Nielsen’s 10 Usability Heuristics & Outlooks for the Future,” in which the 10 Nielsen heuristics were combined with the Schneiderman and Plaisant 8 golden rules to create the Nielsen-Shneiderman’s Heuristics (2017)).

### 8.3 - Cognitive Walkthroughs

A cognitive walkthrough is a method used to evaluate the learnability of a product or service (Wilson 2014). There are many types of cognitions: thinking, memory, learning, decision-making, talking, and seeing. Cognitions can be broken down into further categories: experiential and reflective, as well as fast and slow thinking (Preece, Rogers & Sharp, 2015). Both modes, experiential and reflective cognition, help people decide what they do every day. We use this type of thinking to create a model of how other people process information in their heads. This information can also be used to explore how each individual can be supported by different kinds of technologies. This will help in how developers build a friendly user interface. If developers are able to understand how human cognition functions, users experiences will be improved when interacting with their systems.

The analysis of cognition can help with the design process of the user interface. It can help to differentiate parts of the interface for the user, which allows users to clearly look at the interface and instantly know where a certain link or button might be. The most important method of cognition is sight. The first thing the user will be able to distinguish from an interface is color or size. The most important part of a website, for instance, could be highlighted, bolded, or bigger in text size. This is the process of selecting things to concentrate on, at a point in time, which is called attention (Preece, Rogers, & Sharp, 2015).

A cognitive walkthrough starts off listing out a variety of tasks the user will most likely take when going through a webpage. This will allow a developer to study the usability of the webpage. The tasks are then elaborated further and become processes for developers to make the webpage user friendly. The processes are usually created as lists, but when these processes become too complicated, it can turn into a diagram such as the one below (Jones).



Larger projects results in a more complicated diagram when the development is nearing its end. Most big diagrams can be broken into smaller parts and are interconnected with one another later in the development process. This way it is easier to read the diagrams and to create a system where a certain function can maintain this set of methods and have primary, secondary or tertiary goals.

### 8.4 – Heuristic Evaluations

Heuristic evaluation is the method through which knowledge of typical users is applied to identify usability problems (Preece, Rogers & Sharp, 2015). Heuristic evaluation is about certain elements in a user interface and sorts out these elements in the interface into certain categories. Heuristic evaluation can be related to high-level design principles. Research in cognitive psychology has shown how people tend to use simple heuristics when making decisions. (Gigerenzer et al, 1999). Heuristic evaluation can be viewed as a pyramid, where the broad and common elements of an interface can be seen on the bottom, and as one goes higher up in the pyramid, the fewer and scarcer some elements become in the interface. The more important elements are seen at the top of the pyramid, and must be represented by trying to distinguish them from the other elements in the interface.

There are many benefits as well as disadvantages of using a heuristic evaluation in interface design. The following is a chart that lists out both the advantages and disadvantages of heuristic evaluation (Molich & Nielsen 1994):

Advantages	Disadvantages
Provide quick/relatively inexpensive feedback for designers	Requires knowledge and experience to apply the heuristics effectively.
Obtaining feedback early in the design process	Hiring trained usability experts are sometimes hard to find and expensive
Assigning the correct heuristic can help suggest the	Recommended to find and use multiple experts and

best corrective measures to designers.	aggregate their results.
You can use it together with other usability testing methodologies	Evaluations may identify more minor issues and fewer major issues
Conduct usability testing to further examine potential issues	

Just because a heuristic evaluation can be used for a variety of topics, it does not mean that it makes other testing obsolete. As there are many advantages to heuristic evaluation, there are nevertheless many disadvantages as well. The chart is only a small representation of the whole heuristic, but it provides enough information to understand the pros and cons of using it. Some of these disadvantages can be made up for by using other testing environments. The benefits of actually using real-life experiments work just as well, for instance, having real users test a GUI in its prototype process. A heuristic evaluation should not replace usability testing (Molich & Nielsen 1994).

### 8.5 - Think-Aloud Evaluations

As opposed to heuristic evaluations, think-aloud evaluations are based on the protocol created by Lewis in 1982 aimed at understanding the cognitive processes of the human brain when completing certain tasks. The participants were told and encouraged to express anything they thought, saw, or felt to better understand the design from the user's point of view. The think-aloud protocol dives deeper into how a real user would interact with the system, and it is believed to better reveal design flaws.

In an example of the think-aloud approach being used, a set of nurses was asked to test a web-based communication tool for scheduling their shifts. As part of the assessment, participants were asked to think aloud while completing their tasks, which included four subtasks related to shift management, including posting and searching for a shift, awarding and requesting a shift, searching for a nurse or shift, and viewing a report or schedule. They were then asked 3 post questions to better gauge their experience. The same evaluation of the scheduling tool was also used in the heuristic approach, and although they both had many similarities in the flaws they noticed, it was found that the think-aloud evaluations tended to point out the most severe obstacles to the user's performance of tasks. Users aren't really bothered with a few "off" designs if it means they can still do their jobs fluidly. (Yen & Bekken, 2009)

Having the participants themselves test the design has its positives and negatives. Sometimes users simply aren't paying much attention to a feature, but that doesn't mean the feature isn't important or that it couldn't be improved. However, it is helpful as a start because it solves the most upfront and severe design flaws. Having users navigate through the applications lets developers see what they click on first or never click on at all, how much time they spend completing a task, or how much time they spend just figuring out what to do. From those observations, developers can optimize the application to better accommodate the users, meaning they can navigate the app more quickly and fluidly than before.

## 8.6 - Classic Experiments

The classic experiment is a type of experimental design that relies on statistical analysis. It is meant to be a firm way of conducting an experiment with minimal outside influences that may affect the current product at hand. It is made up of three main characteristics. One is that there must be a group under study so that it can be observed whether a process has any effect on the group and is worth researching at all. The second characteristic is that the participants need to be randomly selected. This is so that the testers aren't swayed by outside decisions such as giving a sicker patient the newer treatment which can skew the data. If the patient dies, it may not have been from the treatment but from the fact they were sicker when they were given it. The third and last deliverable is that both before and after the experiment, there must be a defined outcome measure. Measuring before and after is to ensure that a change has actually occurred during the process. The approach is set out to determine whether or not there has been an improvement in the design/treatment of a thing. Although some of these characteristics are not always feasible in every situation, it is seen as a very valuable approach. (Greeno, 2002)

Classic experiments offer another option to test what really goes on inside the human brain. This type of method doesn't rely on user input but on data collected and interpreted by the researchers. The theory is that one or more factors will produce an effect that can then be measured. There are a couple of main experimental design experiments, and each has their own benefits and setbacks. There is the experimental causal chain design, the measurement of mediation design, and the moderation of process design. The measurement of mediation states a type of analysis that A influences C through a mediating variable referred to as B. This can be shown in many examples, such as threats to self-esteem (A) had caused an increase in discrimination against a stereotypical subject (C) which is mediated through feelings of self-worth. The other process is achieved by manipulating both the mediating variable and the independent variable, from which we can then surmise strong inferences of a causal chain of events. This type of experiment is referred to as the experimental causal chain design. (Spencer, 2005)

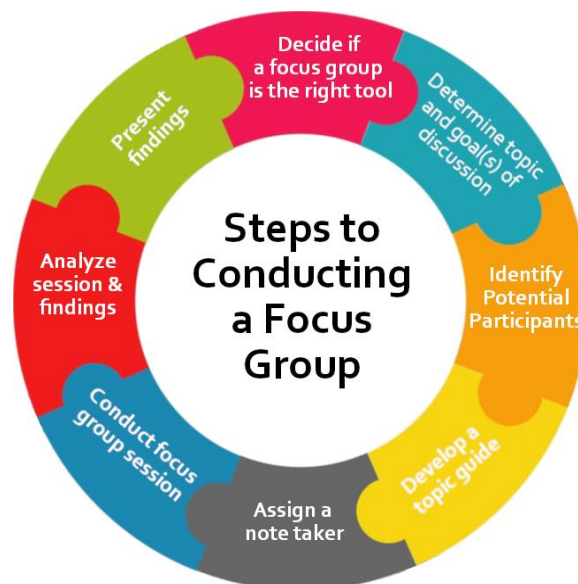
Even though the causal chain design may be seen as the best approach for uncovering the underlying truth of a matter, it isn't perfect. The main setback of the causal chain of events is that it's not always convenient to measure the effect A has on C which is explained by B. Also manipulating or measuring is not always as easy as we would like it to be, in which case the experimental causal chain design is least likely to work and a different approach should be taken. If the design is easy to manipulate but difficult to measure, then this means that it is an unconscious process within the user's head, as there are no reliable or valid measures of construct. So in this case, the moderation of process design would be the best option. And each experiment performs best under certain circumstances which can be found in the graph below. Although it's found that if both ease of manipulating the proposed process and the ease of measuring the proposed process are increasingly difficult, then it doesn't matter which design you use as it isn't likely to produce convincing data or be worth the time and effort. (Spencer, 2005)

## Recomendations for Experimental Designs based on Ease of Manipulating and Measuring

		Ease of manipulating proposed process	
		Easy	Hard
Ease of measuring proposed process	Easy	Experimental casual chain design	Measurement of mediation design
	Hard	Moderation of process design	No design is likely to work

### 8.7 – Focus Groups

Focus groups are a specific form of group interviews that rely on open and quality communication between research participants that produces and generates data, where group interaction is beneficial, even crucial, to the method of data gathering and evaluation. The advantage of using a focus group is that it explores both what people think as well as how and why they think; focus groups act as a conversation between participants, each responding not only to the interviewer/researcher's questions but also to one another, exchanging information and personal experiences. (Kitzinger, 1995, p.299)



As presented in the above figure, there exists a recommended approach to utilizing and conducting a study involving a focus group; the above step-by-step method is a generalized guide to properly employing a focus group study and maximizing results (Stewart & Shamdasani, 2014). First and foremost, it's important to consider whether a focus group is the right tool for the job or if another method might be more appropriate. In the case of user interactive design, a focus group is particularly handy when seeking initial requirements,

envisioned designs, and feedback from the client. But it's important to determine before beginning what the primary focus and expected results of the group study are so that the interviewer can tailor and moderate discussion. Potential participants of the study must be identified; again, in the case of interactive design, the client and end-users will be the target audience and study participants. After these steps are met, designing and developing a structured guide will aid the interviewer in keeping on track; similarly, assigning a note taker allows the interviewer to focus on communication while feedback and discussion is noted elsewhere. In addition to a note taker, recording the group by video or audio alone is also recommended. Finally, the focus group session can be conducted, followed by an analysis of the gathered data which can then be presented or used to make critical decisions moving forward. (Morgan, 1996)

Focus groups serve three distinct purposes: self-contained methods, supplementary sources, and multimethod studies. Focus groups employed as part of an interaction design project might fall under any one of these three purposes. As a self-contained method, a focus group would be the only method of gathering data. While this may not be the ideal scenario, it is certainly plausible and may be optimal in certain circumstances. But a focus group might only serve as a supplementary source of data, where another method takes precedence and acts as the primary source of data collection. Lastly, a multimethod study would employ a focus group method alongside multiple other methods, such as a survey or other means. This method allows for more cross-section analysis of gathered data and can prove very useful to narrowing down results. (Morgan, 1996)

Focus groups originated as a way to gauge audience reactions to film and television; however, as the value of the method was recognized, its reach broadened and it became a staple of qualitative research designs. Focus groups are equally valuable when it comes to evaluating user interfaces and user interaction with design. A study conducted among elders transitioning to digital calendars utilized a study group method to better understand what challenges the seniors were facing. On the surface, it appeared that the interface was too complicated or user-hostile; however, after discussing the issue during several focus groups, the actual issues rose to the surface—it wasn't that the interface was too difficult for the seniors to handle, but that they found it challenging to make a daily habit and develop a routine involving the app. (Baric, Andreassen, Ohman, & Hemmingsson, 2019)

Focus groups can play an important role in various steps of the design process. Before building or even designing the interface, a group discussion can help focus the goal of the project. What are the users looking for? How accessible do they imagine the program to be? Who is the program being built for, and how user-friendly does it need to be? What issues and requirements do the stakeholders bring to the table? What quality-of-life improvements can be derived from the wishes of the intended user group? As individuals are able to discuss as a group, they're able to bounce ideas off one another and develop a vision as a unit. User feedback is again crucial once the project reaches an evaluation point; getting direct input from the user group or the client(s) is critical to delivering a well-formed end product. Users can catch bugs the designer may not have seen themselves, users can give honest input on design choices, and users going in blind can offer feedback on accessibility and usability. (Abrams, Maloney-Krichmar, & Preece, 2004)

Even after the product is completed and released, evaluation should be continued; here again, a focus group can provide quality data about user satisfaction, design, and functionality (Abrams et al., 2004). As quoted by Patricia Sullivan, "studies conducted during development serve as [a] litmus for developers, and thus steer aspects of the development" (1991, p.106). That is, any study conducted among the user base or client(s) during (and after) development helps guide the developer towards a clearer goal. User feedback will direct



the development process closer towards or further away from different design elements as response and opinions are taken into consideration.

Perhaps one of the greatest benefits of utilizing the focus group methodology is that it takes into account a range of responses and opinions rather than solely individual ones. When designing an interface that's to be taken advantage of by a wide breadth of users, it's important to cater to the whole rather than the one. The hallmark of focus group research is its "explicit use of the group interaction to product data and insights that would be less accessible without the interaction." (Morgan, 1996)

## **8.8 - Concise Summary**

Interfaces are used every day by people to interact with the world around them, which is why they should be properly tested. Of the different types of interface evaluations, there are three general categories: controlled settings with users, natural settings with users, and settings not involving users. Each have their own benefits, but among all interfaces there are guidelines and standards which should be sought after. Some include efficiency of use, consistency, error handling and freedom of control, which are referred to in Jakob Nielsen's 10 heuristics and Shneiderman and Plaisanrt's eight golden rules. Within the three main categories there are many different methods of testing.

Cognitive walkthrough is the evaluation of a good or service. The way it works is to think about how other users who would use the product try to process on how to make it function. It helps developers to create a simplistic and easy-to-use product in order for a better user experience. It also helps with errors; if an error occurs, the system needs to explain to the user how to fix the error.

Heuristic evaluation is used to establish a clear line between different elements of a user interface. This can be represented as a pyramid, where the bottom includes the most common elements and the top, the more important elements. This helps with separating the least important elements from the most important elements. This allows developers to focus more on different elements that are inserted within an interface.

The thinking-aloud protocol is another method of testing usability. It is a method that assigns certain tasks for users to complete and encourages them to verbalize their thoughts while completing the task, to gain a better understanding of the way they perceive and interact with the interface.

Classic experiments are a subtype of the experimental design, where an experiment is conducted with minimal outside influences to better determine the cause of something. They are derived from a simple hypothesis, and that hypothesis is then tested under certain circumstances to determine if it was correct.

Focus groups serve as a form of group interview that allows for fluid communication and feedback. In research, this qualitative method of study provides direct, open, and honest reporting from subjects. In the field of user interface and design interaction, focus groups serve well as a means of gathering feedback from clients and potential user bases. Focus groups capitalize on group interaction to pull data that otherwise might not be attainable.

## 8.9 - Extended Resources

### Descriptions & Links

1. Youtube video discusses the thinking aloud methodology for website usability testing  
<https://www.youtube.com/watch?v=qQxq5Muo5GI>

---

2. Research methods of the experimental design, describing the steps that bring about an experimental conclusion.  
<https://www.youtube.com/watch?v=qtLnBz6lbRQ>

---

3. This Ted-Ed video explains how focus groups works, how they came about, and the uses they serve in research.  
<https://www.youtube.com/watch?v=3TwgVQIZPsw>

---

4. This lecture video from Yale University discusses different methods of qualitative research, including focus groups and their place in research methodologies.  
<https://www.youtube.com/watch?v=cCAPz14yjd4>

---

5. This video discusses Jakob Nielsen's 10 heuristics to better usability  
[https://www.youtube.com/watch?v=\\_RxfU6dPZuU](https://www.youtube.com/watch?v=_RxfU6dPZuU)

---

6. Short youtube video explains the definition of Cognitive Walkthrough and gives examples on how it used in a work environment.  
<https://www.youtube.com/watch?v=a3niUbFJIEw>

---

## 8.10 - References

- Abras, C., Maloney-Krichmar, D., & Preece, J. (2004). User-centered design. *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications*, 37(4), 445-456.
- Baric, V., Andreassen, M., Öhman, A., & Hemmingsson, H. (2019). Using an interactive digital calendar with mobile phone reminders by senior people - a focus group study. *BMC Geriatrics*, 19(1), 116. <https://doi-org.proxy.kennesaw.edu/10.1186/s12877-019-1128-9>
- Blackmon, M. H. Polson, P.G. Muneo, K & Lewis, C. (2002) Cognitive Walkthrough for the Web CHI 2002 vol.4 No.1 pp463–470. <https://www.interaction-design.org/literature/article/how-to-conduct-a-cognitive-walkthrough>
- Charfi, S., Ezzedine, H., & Kolski, C. (2013). RITA: A framework based on multi-evaluation techniques for user interface evaluation: Application to a transport network supervision system. *2013 International Conference on Advanced Logistics and Transport*. doi:10.1109/icadlt.2013.6568470
- Haklay, M., & Tobón, C. (2003). Usability evaluation and PPGIS: towards a user-centred design approach. *International Journal of Geographical Information Science*, 17(6), 577. <https://doi-org.proxy.kennesaw.edu/10.1080/1365881031000114107>
- Gigerenzer, G. (2008) *Gut Feelings*. Penguin.
- Gonzalez-Holland, E., Whitmer, D., Moralez, L., & Mouloua, M. (n.d.). Examination of the use of Nielsen's 10 usability heuristics & outlooks for the future. *Proceedings of the Human Factors and Ergonomics Society, 2017–October*, 1472–1475. <https://doi-org.proxy.kennesaw.edu/10.1177/1541931213601853>
- Greeno, C. (2002). Major Alternatives to the Class Experimental Design. 733-734
- Lewis, C., Wharton, C. (1997). *Handbook of Human-Computer Interaction (Second Edition)*. 717-732. <https://doi.org/10.1016/B978-044481862-1.50096-0>
- Molich, R. and Nielsen, J., *Improving a human- computer dialogue*, *Communications of the ACM*, 33(3), 338-348, (1990).
- Morgan, D. L. (1996). *Focus groups as qualitative research*(Vol. 16). Sage publications.
- Nielsen, J., *Enhancing the explanatory power of usability heuristics*, CHI'94 Conference Proceedings, (1994). <https://www.usability.gov/how-to-and-tools/methods/heuristic-evaluation.html>
- Nielsen, J. (1994). Enhancing the explanatory power of usability heuristics. In Proceedings of the SIGCHI conference on Human Factors in Computing Systems, 152-158.
- Preece, J., Rogers, Y., & Sharp, H. (2015). *Interaction design: Beyond human-computer interaction*.
- Schneiderman, B. & Plaisant, C. (1998). Designing the user interface, 3rd ed. Reading: Addison-Wesley.

- Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a Causal Chain: Why Experiments Are Often More Effective Than Mediational Analyses in Examining Psychological Processes.
- Stewart, D. W., & Shamdasani, P. N. (2014). *Focus groups: Theory and practice* (Vol. 20). Sage publications.
- Sullivan, P. (1991, October). Multiple methods and the usability of interface prototypes: the complementarity of laboratory observation and focus groups. In *Proceedings of the 9th annual international conference on Systems documentation* (pp. 106-112). ACM.
- Wilson, C. (2014). *User Interface Inspection Methods: A User-Centered Design Method*. 65-79. <https://doi.org/10.1016/B978-0-12-410391-7.00004-X>
- Yen, P., Bakken, S. (2009). A Comparison of Usability Evaluation Methods: Heuristic Evaluation versus End-User Think-Aloud Protocol – An Example from a Web-based Communication Tool for Nurse Scheduling. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2815403/>

Hero Image: Author/Copyright holder: Nicola Jones. Copyright terms and licence: CC BY-ND 2.0

